Journal of Computing and Applied Informatics (JoCAI) Vol. 01, No. 1, 2017 | 38 - 44

talenta PUBLISHER

DATA SCIENCE

Journal of Computing and Applied Informatics

Implementation and Comparison of Berry-Ravindran and Zhu-Takaoka Exact String Matching Algorithms in Indonesian-Batak Toba Dictionary

Efelin O. Siburian[1], Mohammad Andri Budiman[2], and Jos Timanta Tarigan[3]

[1,2,3]Department of Computer Science, Universitas Sumatera Utara, Medan, Indonesia

Abstract. Indonesia has a variety of local languages, which is the Batak Toba language. This time, there are still some Batak Toba people who do not know speak Batak Toba language fluently. Nowadays, desktop based dictionary is one of reference that very efficiently used to learn a language and also to increase vocabulary. In making the dictionary application, string matching can be implemented for word-searching process. String matching have some algorithm, which is Berry – Ravindran algorithm and Zhu-Takaoka algorithm and will be implemented on the dictionary application. Zhu-Takaoka algorithm and Berry – Ravindran algorithm have two phases, which are the preprocessing phase and the searching phase. Preprocessing phase is a process to make the shifting values according to in pattern that input by user. To know the shifting value with Zhu-Takaoka algorithm, it's need Zhu-Takaoka Bad Character (Ztbc) and Boyer-Moore Good Suffix (Bmgs). Then, Ztbc will be compared to Bmgs to get the maximum value of them that will be set as shifting value. While Berry-Ravindran algorithm, to know the shifting value is needed Berry-Ravindran Bad Character, which the two characters right of the text at the position m + 1 and m+ 2, is needed to determine the shifting value, where m is length of the pattern.

Keyword: Algorithm, String Matching, Zhu-Takaoka, Berry-Ravindran, Dictionary.

Abstrak. *Indonesia memiliki beragam bahasa daerah, salah satunya adalah Bahasa Batak Toba. Saat ini masih banyak masyarakat bersuku Batak Toba yang belum fasih dalam berbahasa Batak Toba. Kamus dapat dijadikan sebagai salah satu sarana untuk belajar berbahasa. Dalam pembuatan aplikasi kamus, string matching dapat diimplementasikan dalam proses pencarian katanya. Ada beberapa algortima dalam String matching antara lain algoritma Zhu-Takaoka dan algoritma Berry-Ravindran dan akan diimplementasikan pada aplikasi kamus tersebut. Algoritma Zhu-Takaoka dan algoritma Berry-Ravindran memiliki dua fase yaitu fase preprocessing dan fase pencarian. Fase preprocessing merupakan proses untuk mendapatkan nilai pergeseran sesuai dengan pattern yang dimasukkan. Nilai pergeseran ditentukan dari aturan algoritma Zhu-Takaoka dan algoritma Berry-Ravindran. Untuk mengetahui nilai pergeseran dengan algoritma Zhu-Takaoka diperlukan Zhu-Takaoka Bad Character dan Boyer-Moore Good Suffix kemudian keduanya akan dibandingkan untuk mencari nilai terbesar yang akan dijadikan sebagai nilai pergeseran. Sedangkan pada algoritma Berry-Ravindran untuk mengetahui nilai pergeseran*

*Corresponding author at: Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Jalan Alumni No. 9 Kampus USU, Medan 20155, Indonesia

E-mail address: efelin.o.siburian@students.usu.ac.id (Efelin O. Siburian), mandrib@usu.ac.id (Mohammad Andri Budiman), jostarigan@usu.ac.id (Jos Timanta Tarigan)

Copyright © 2017 Published by Talenta Publisher
ISSN: 2580-6769 | e-ISSN: 2580-829X | DOI: 10.32734/jocai.v1.i1-67
Journal Homepage: https://talenta.usu.ac.id/JoCAI

diperlukan Berry-Ravindran Bad Character yang merupakan dua karakter sebelah kanan teks pada posisi m+1 dan m+2 dimana m merupakan panjang pattern.

Kata Kunci: *Algoritma*, String Matching, Zhu-Takaoka, Berry-Ravindran, *Kamus*.

Received 01 May 2017 | Revised 02 June 2017 | Accepted 05 July 2017

1. Introduction

Batak ethnic is one of a big nation ethnic in Indonesia. But there are so many Batak people who do not know speak batak language well. So that, the writer give a solution to defend Batak language. The writer build a dictionary application that can translate Bahasa Indonesia in to Batak language and vice versa by implementing string matching algorithm that is Berry-Ravindran algorithm and Zhu-Takaoka algorithm with desktop base.

Algorithm can be defined as a computation process who take or determine some of value as input and produce or determine some of value value as output. Or algorithm is sequence of computation steps that change an input become an output [2].

2. Method

Berry-Ravindran Algorithm is a string matching algorithm which is a blend between Quick Search algorithm and Zhu-Takaoka algorithm. This algorithm is proposed by T. Berry and S. Ravindran in 1999. This algorithm do the shift with compute the shifting of bad character which it's value obtained form preprocessing phase. Berry-Ravindran algorithm do the string matching from left to right.

BM‟ (Zhu-Takaoka) algorithm is modification of Boyer Moore algorithm that have the same characteristic in string searching process. The characteristic is consists of two phases were preprocessing phase and searching phase. The difference of Boyer-Moore algorithm and Zhu-Takaoka algorithm is on bad character rule determine phase. In Boyer-Moore algorithm, bad character just one dimension butin Zhu-Takaoka modified become array of two dimensions. Zhu-Takaoka algorithm do string matching from right to left. In preprocessing phase, Zhu-Takaoka algorithm build bad character table with two dimension because this algorithm doing computation for pair of characters. Flowchart of Berry-Ravindran and Zhu-Takaoka searching process can be seen in Figure 1 and 2 below.

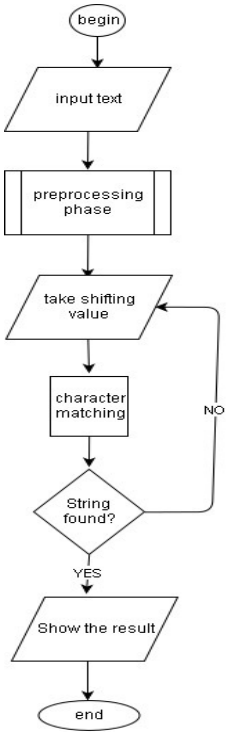

Figure 1. Flowchart of Berry-Ravindran Searching Process

3. Result And Discussion

A. Berry-Ravindran Algorithm

Before doing searching process, Berry-Ravindran algorithm have preprocessing phase to determine the shifting value. Figure 2 below show the shifting value of pattern 'aha'. And the sample text in database are agat,

brBc	A	H	*
A	1	1	1
H	2	5	5
*	4	5	5

Figure 2. The Shifting Value of Pattern 'Aha'

In the first searching process, pattern 'aha' be matched with text 'agat'. If the shifting value is bigger than difference long of pattern and text, so system will not doing searching process and it will be return zero value its mean that pattern not found. The difference long of 'agat' and 'aha' is one, it's mean that searching process just have probability once shifting. In Figure 3, text 'agat' having the addition of two characters '00' that aims to avoid ArrayIndexOutOfBound. To know the shifting value, Berry-Ravindran take two caharacters right of the text. In the case of 'agat' it just have one character right that is T, so with the addition of two characters we can get the shifting value.

A	G	A	T	0	0
A	H	A			
brBc [T][0] = 5					

Figure 3. Berry-Ravindran Searching for text 'agat'

We can see that the shifting value is 5. It is bigger than difference long of pattern and text, so that system will be return zero value that means 'aha' not found in 'agat' and it will be taking the next text to be matched.

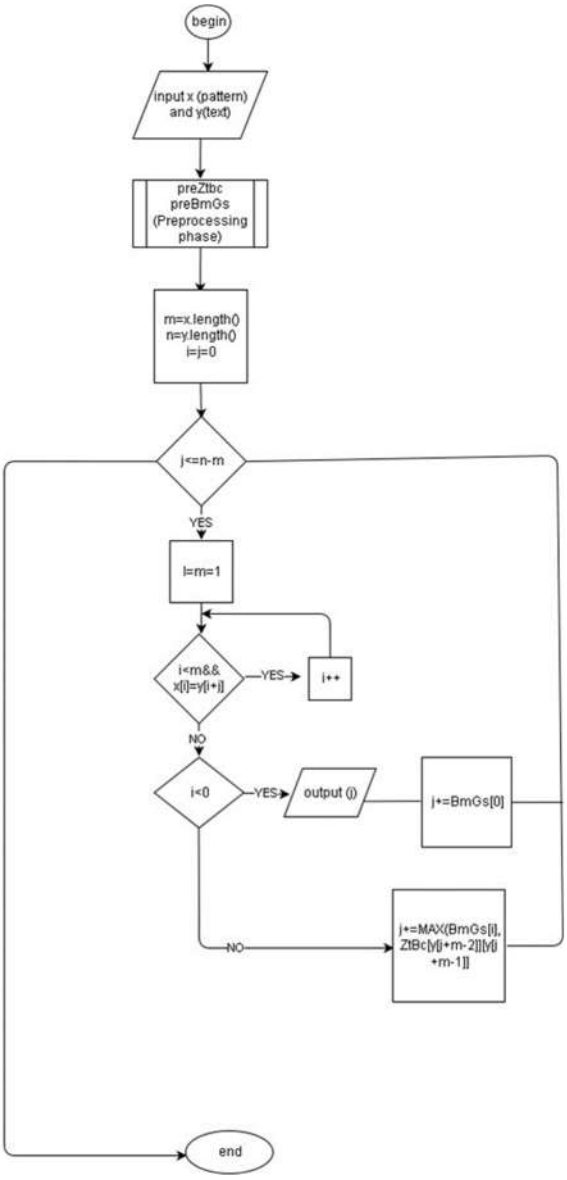

Figure 4. Flowchart of Zhu-Takaoka Algorithm Searching Process

B. Zhu-Takaoka Algorithm

The first step is preprocessing phase that making two shifting tables, ZtBc (Zhu-Takaoka Bad Character) and BmGs (Boyer-Moore Good Suffix). The result of preprocessing of pattern 'aha' can be seen in Figure 5 and 6 below.

ZtBc	**A**	**H**	*
A	2	1	3
H	2	3	3
*	2	3	3

Figure 5. Zhu-Takaoka Bad Character Table

I	**0**	**1**	**2**
x[i]	A	H	A
suff[i]	1	0	3
bmGs[i]	1	2	1

Figure 6. Boyer-Moore Good Suffixes Table

Steps of the searching of pattern 'aha' in text 'mahap' with Zhu-Takaoka algorithm can be seen below.

Step 1

Window		A	H		
Text	M	A	H	A	P
Pattern	A	H	A		
I	0	1	2		

Figure 7. Step one of searching in the text

Ztbc [A][H]=1

Bmgs[2]=1

Bmgs [2] is equal with Ztbc [A][H] then do one-time shifting

Step 2

Window			H	A	
Text	M	A	H	A	P
Pattern		A	H	A	
I		0	1	2	

Figure 8. Step two of searching in the text

Characters is matched.

Shift do as much bmGs[0] = 1

Step 3

Window				A	P
Text	M	A	H	A	P
Pattern			A	H	A
I			0	1	2

Figure 9. Step two of searching in the text

ztBc[A][P] = 3

bmGs[i] = bmGs[2] = 1

do three times shifting

Cause of the length of the text already exhausted, so matching process terminated. From the example above, it can be concluded that text 'mahap' and pattern 'aha' produce a pattern that matched by using Zhu-Takaoka algorithm. -Takaoka produce 1 pattern that matching.

4. Conclusion

The conclusion of this research as follows:

1. The application that is created is Indonesian-Batak Toba dictionary desktop base by using Berry-Ravindran and Zhu-Takaoka algorithm.
2. Berry – Ravindran and Zhu – Takaoka algorithm can be implemented on Indonesian-Batak Toba dictionary application and it can be run well.

This research showed where the longer the text character, then the shorter the running time or in other words, running time and length of text characters is inversely proportional,. Based on system testing, it showed that the time for the search processing required by Zhu-Takaoka algorithm is shorter than Berry-Ravindran algorithm.

REFERENCES

[1] Charras. C. & Lecroq. T. 2004. Handbook of Exact String Matching Algorithms (E Book). English. College Publications.

[2] Cormen, T.H., Leiserson, C.E., Rivest, R.L., & Stein, C.2001. Introduction to Algorithms. 2nd Edition. The MIT Press : London.

[3] Hussain, Iftikhar., Ali, Imran., Zubair, M. & Bibi, N. 2010. Fastest Approach to Exact Pattern Matching. Proceedings of 2010 International Conference on Information and Emerging Technologies (ICIET), pp. 1 -5.

[4] Limbong, Bernhard.2014. Kamus Bahasa Batak Toba-Bahasa Indonesia. Permata Aksara : Jakarta.

[5] Michailidis. P. D., Margiritis. G. K. 2009. Experimental Study on Variants of the Zhu-Takaoka String Matching Algorithm. University of Macedonia: Thessaloniki.

[6] Pressman, R. S. 2010. Software Engineering: A Practitioner's Approach. 7th Edition. McGraw-Hill: New York.

[7] Purba, J. I. 2016. Implementasi Pencocokan String Menggunakan Algoritma Berry Ravindran pada Aplikasi Kamus Bahasa Indonesia –Simalungun Berbasis Android.Skripsi.Universitas Sumatera Utara.

[8] Ramdhani, P. P. 2012. Analisis Perbandingan Performansi Algoritma Zhu-Takaoka dan Karp-Rabin pada Pencarian Kata di Rumah Buku Baca Sunda. Skripsi. Universitas Komputer Indonesia.

[9] Syuhada,F.2016. Implementasi Algoritma Zhu-Takaoka pada aplikasi Terjemahan Al Quran Berbasis Android.Skripsi.Universitas Sumatera Utara.

[10] Wicaksono. A. K. 2015. Perbandingan Algoritma Horspool dan Algoritma Zhu-Takaoka Dalam Pencarian String Berbasis Desktop. Skripsi. Universitas Multimedia Nusantara.

[11] Whitten, J.L., Bentley, L.D. & Dittman, K.C. 2004. Metode Desain & Analisis Sistem. Terjemahan TIM Penerjemah ANDI. ANDI : Yogyakarta.